THE STRUCTURE OF SYNTHETIC POLYPEPTIDES

BY LINUS PAULING AND ROBERT B. COREY

GATES AND CRELLIN LABORATORIES OF CHEMISTRY,* CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA, CALIFORNIA

Communicated March 31, 1951

In a preliminary communication last year[1] we stated that there are only two helical configurations of polypeptide chains in which the residues are all equivalent and intramolecular hydrogen bonds are formed, and in which the interatomic distances, bond angles, and other structural features, especially the coplanarity of the conjugated amide system, are as required by earlier work in these Laboratories on amino acids, simple peptides, and other substances related to proteins. These two helical configurations were described in detail in a later paper[2] and it was mentioned that there is evidence that they occur in α keratin, α myosin, supercontracted keratin and myosin, and other fibrous proteins, and also constitute an important structural feature of hemoglobin and other globular proteins.[3] In the following paragraphs we discuss evidence that one of the helical structures is assumed also by synthetic polypeptides.

Oriented films and fibers of several synthetic polypeptides have been prepared and examined by x-ray diffraction by Bamford, Hanby, and Happey,[4] and the oriented films have been investigated with polarized infrared spectroscopy by Ambrose and Elliott.[5] The best photographs were given by poly-γ-methyl-L-glutamate and poly-γ-benzyl-L-glutamate. The x-ray data indicate an identity distance in the direction of the fibers of 5.50 A for the first substance, and a slightly larger value for the second substance. The dichroism observed for the N—H stretching and bending infrared absorption bands and for the C=O stretching band indicates that these groups are oriented nearly parallel to the fiber axis, and the conclusion is accordingly drawn that intramolecular hydrogen bonds are formed nearly parallel to this axis. These authors have interpreted all of their data as providing strong support for the α_{II} structure shown in figure 1. This structure was first proposed by Huggins[6] and has been discussed by Zahn,[7] Simanouti and Mizushima,[8] and Ambrose and Hanby.[9]

The α_{II} structure must, however, be rejected. In the structure as discussed by Bamford and coworkers the amide group is not assigned a planar configuration. There is, in fact, very strong theoretical and experimental evidence that the carbon-nitrogen bond has a large amount of double-bond character, and that the four atoms adjacent to these two atoms must form a coplanar system with them. The carbon-nitrogen distance is 1.32 A, which is 0.15 A less than the single-bond distance between these atoms. This shortening corresponds to about 50 per cent double-bond character,

which is great enough to lead to effective coplanarity of the system. (Ambrose and Elliott[5] say "It remains to be shown whether resonance in polypeptides and proteins effectively limits the rotation about the C—N bond, and we regard the matter as by no means settled." In fact, however, our present knowledge of structural chemistry is such that there can be no doubt on this point: non-planar configurations of this group must surely be accompanied by pronounced instability.) In our investigation of helical configurations of the polypeptide chain[2] we found it impossible to construct an acceptable configuration resembling α_{II}. The closest approximation to an acceptable configuration that can be constructed, in which the hydrogen atom of the N—H group is brought to within 1.8 A of the carbonyl oxygen atom, is unsatisfactory because the N—H···O angle differs from a straight angle by about 70°, and thus this configuration would be presumed

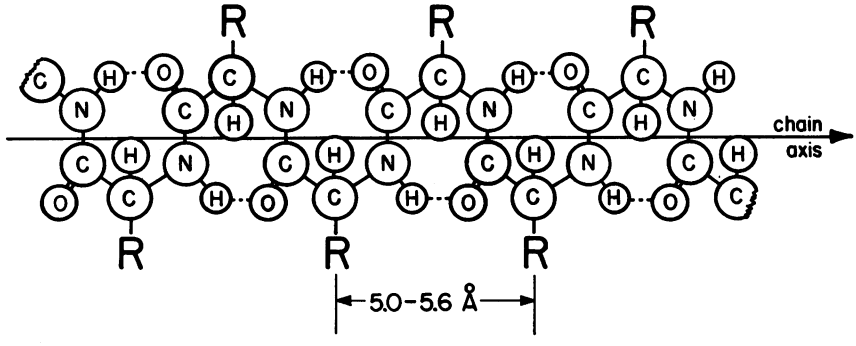

FIGURE 1

The α_{II} structure for the polypeptide chain, discussed by Bamford, Hanby, and Happey, and by Ambrose and Elliott.

not to correspond to a reasonably strong hydrogen bond. It was for this reason that the α_{II} structure was eliminated in our earlier considerations. There are also several other arguments against the α_{II} structure. Ambrose and Elliott mention that their model gives a dichroic ratio of 1.41:1 for the C=O stretching calculated for perfectly oriented molecules, and that the observed dichroism in poly-γ-benzyl-L-glutamate, 2.6:1, is considerably greater than this value. Moreover, the α_{II} structure corresponds to an extension of only approximately 40 per cent during the $\alpha \rightarrow \beta$ transformation. Ambrose and Elliott suggest that the observed reversible extension of wool, 100 per cent or more, is not to be interpreted as necessarily requiring a similar reversible extension from the α to the β configuration of the polypeptide chain, because there is present some amorphous material in wool; but it seems to us, as also to Astbury,[10] very unlikely that such a great discrepancy could exist. In addition, it may be pointed out that the

synthetic polypeptides seem to be hexagonal or closely pseudohexagonal in structure, and the flat α_{II} configuration provides no explanation of this fact.

All of these difficulties are overcome if it is assumed that the synthetic polypeptides have the configuration of the third-amide hydrogen-bonded helix which we have described. In this helix the distance per residue along the helical axis is predicted to be approximately 1.50 A and there are approximately 3.7 residues per turn. The translation along the axis per turn is thus predicted to be about 5.5 A. This is in excellent agreement with the fiber axis translations reported by Bamford, Hanby, and Happey,

TABLE 1

X-RAY DATA FOR POLY-γ-METHYL-L-GLUTAMATE
HEXAGONAL UNIT WITH a_0 = 11.96 A, c_0 = 27.5 A

Equator:

$HI \cdot L$	$d_{obs.}$	$d_{calc.}$	F_1[a]	F_2	$I_{2_{calc.}}$	$I_{obs.}$
10·0	10.35 A	10.35 A	18.5	16.2	360	VVS
11·0	5.98	5.98	5.1	6.2	29	S
20·0	5.22	5.18	1.3	3.2	7	M
21·0	3.89	3.91	−4.0	−4.8	22	S
30·0	3.45	3.45	−5.0	−7.1	21	S
22·0	3.00	2.99	−5.0	−6.1	14	S
31·0	2.87	2.87	−4.9	−1.8	2	W

Second hyperbola

$HI \cdot L$	$d_{obs.}$	$d_{calc.}$
11·2	5.51 A	5.48 A
20·2	4.88	4.87

Fifth hyperbola

$HI \cdot L$	$d_{obs.}$	$d_{calc.}$
10·5	4.82 A	4.86 A
11·5	4.05	4.05
20·5	3.75	3.77

Sixth hyperbola: a polar arc with $d_{obs.}$ = 4.43 A; $d_{calc.}$ for {10.6}, 4.20 A.

[a] Calculated for cylindrically symmetrical distributions of atomic centers, with radii from helical axis 2.29 A for C, 1.59 A for N, 1.61 A for C′, 1.74 A for O, and 3.34 A for βC.

5.50 A for poly-γ-methyl-L-glutamate and slightly larger for the benzyl ester. Moreover, the helical molecules would be expected to pack together essentially as would cylinders, in a hexagonal packing, and the predicted interplanar distances for the hexagonal lattice agree well with those observed, and shown in tables 1 and 2.

Bamford, Hanby, and Happey interpreted the data for the benzyl ester in terms of an orthorhombic unit with axes 10.35 A, 5.98 A, and 5.50 A. The number of molecules in this unit calculated from the density 1.34 is 1.9. The authors note that the first two axial lengths are in the ratio $\sqrt{3}:1$, which is compatible with hexagonal symmetry, and they ask to what extent the x-ray results are consistent with a helical configuration of the mole-

cules. They point out that a hexagonal unit with a_0 = 11.96 A and c_0 = 5.50 A and containing three amino-acid residues, corresponding to a three-fold helix, would have a density 1.06 g cm^{-3}, which is too small. Our helix, with about 3.7 residues per turn, would lead, however, to the density 1.29, which is acceptable. A similarly satisfactory value of the density, 1.3, is also calculated for the corresponding hexagonal unit for the benzyl ester, with interplanar distances given in table 2.

Although most of the reflections observed by Bamford, Hanby, and Happey are accounted for by the orthorhombic unit or the hexagonal unit

TABLE 2

X-RAY DATA FOR POLY-γ-BENZYL-L-GLUTAMATE

PSEUDO-ORTHORHOMBIC UNIT WITH a_0 = 25.0 A, b_0 = 17.3 A, c_0 = 14.42 A

Equator:

hkl	$d_{calc.}$	$d_{obs.}$	F_1	F_2	$I_{2_{calc.}}$	$I_{obs.}$
100	25.0	...	0	1.3	0.5	W
001	14.4	...	0	1.1	0.4	W
101, 200	12.5	12.6	21.5	17.8	165	VVS
002, 301	7.21	7.21	9.5	11.1	35	VS
202, 400	6.25	6.21	6.0	9.0	20	S
103, 402, 501	4.72	4.87	−0.8	−2.2	2	M (diffuse)
303, 600	4.17	4.07	−3.8	−7.4	9	S (diffuse)
004, 602	3.61	3.85	−4.4	−4.8	4	
204, 503, 701	3.49	...	−4.8	+0.8	0.2	...

First hyperbola: faint reflections at 16 A.
Second hyperbola: faint reflections at 8 A.
Third hyperbola:

hkl	$d_{calc.}$	$d_{obs.}$	$I_{obs.}$
151, 250	5.24 A	5.27 A	VS
052, 351	4.50	4.48	S (diffuse)
252, 450	4.24	4.08	
153, 452, 551	3.65	3.60	

with c_0 = 5.50 A, some reflections which they could not explain were reported by them. They concluded that these reflections, observed for both of the esters, are due to a second phase in each specimen, because the reflections seemed not to have any obvious relation to those that had been indexed. The second phase would have to be assumed to have parallel orientation to the main phase. It has seemed to us likely that these weak reflections indicate the presence of a larger unit for each crystal. A larger unit of structure would, of course, be predicted on the basis of our 3.7-residue helix, inasmuch as a true identity distance along the helical axis could occur only after a number of turns. Even though the interatomic interactions that are operative in stabilizing the helix would not be expected,

in a non-crystalline phase, to lead to a rational number of residues per turn, the forces operating during crystallization might well be great enough to produce a small torque in the helix, such as to cause it to assume a configuration with a rational number of residues per turn. The three simplest rational helixes of the 3.7-residue hydrogen-bonded type are those with 11 residues in 3 turns (3.67 residues per turn), 15 residues in 4 turns (3.75 residues per turn), and 18 residues in 5 turns (3.60 residues per turn). We have found that the weak reflections for the methyl ester can be accounted for on the basis of the 18-residue 5-turn helix, and those for the benzyl ester on the basis of the 11-residue 3-turn helix. Plans of these helixes are shown as figures 2 and 3.

The 18-residue 5-turn helix has a sixfold screw axis of symmetry, and it would accordingly be expected that these helical molecules would arrange themselves side by side in hexagonal packing in such a way that the unit

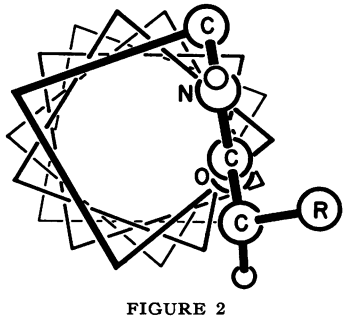

FIGURE 2

Plan of the 18-residue 5-turn helix.

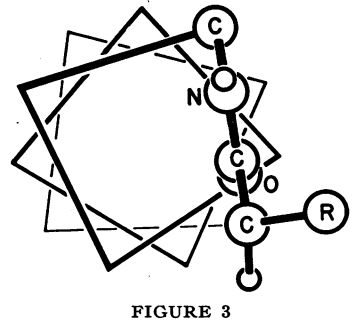

FIGURE 3

Plan of the 11-residue 3-turn helix.

of structure would contain only one 18-residue segment of one helix. The predicted dimensions of the hexagonal unit of structure would, then, be $a_0 = 11.96$ A, $c_0 = 27.5$ A. All of the data reported by Bamford, Hanby, and Happey are accounted for by this unit, as shown in table 1. The two reflections in the second hyperbola and the one in the sixth hyperbola are those which are not accounted for by the smaller unit. The earlier authors described the two reflections in the second hyperbola as corresponding to a c-axial length of about 13.3 A, which is close to half of our value of c_0. The reflection at 4.43 A is a polar arc, which probably could not be measured very accurately.

The determination of the structure of the crystals of the methyl ester would involve only the determination of the orientation of one of the helixes and of the positions of the atoms in the side chains. Simple considerations show that the observed intensities are approximately those predicted for a structure of this sort.

We have calculated intensities of the equatorial reflections for this hexagonal structure, with the four peptide atoms CNCO and the β carbon atom distributed over cylindrical surfaces, the structure factor then being $F_{HI\cdot L} = \sum_i f_i J_0(2\pi\rho_i/d_{HI\cdot L})$; here f_i is the atomic structure factor for the ith atom, ρ_i is the distance of the atom from the helical axis, as given by our calculations,[2] and J_0 is the Bessel function of order zero. The values found are included in table 1, as F_1. It is seen that there is rough general agreement. Our calculations have included only five of the ten heavy atoms per residue; and although it may be predicted that the contribution of the

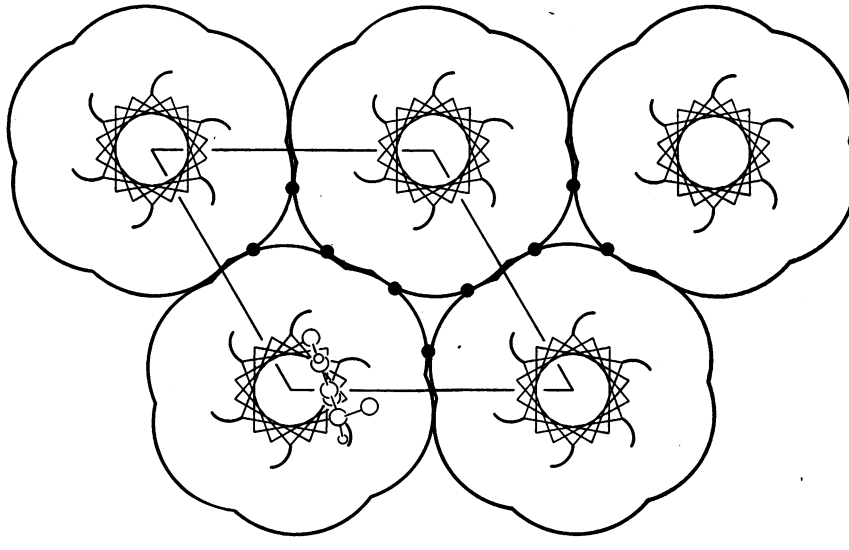

FIGURE 4

Proposed structure of poly-γ-methyl-L-glutamate.

other five (side-chain) atoms would be less, because of greater mutual interference, than that of the five main-chain atoms, it should be significant, and can explain the observed medium intensity of $\{20.0\}$ and weak intensity of $\{31.0\}$. That this can be done is seen by comparison with F_2, and with I_2, the corresponding intensities $pLPF_2^2$, in which p is the frequency factor, L the Lorentz factor, and P the polarization factor. In this calculation it is assumed that side-chain atoms are grouped about the 3-fold screw axes, in the positions xyz, with $xy = {}^3/_7\,{}^6/_7, {}^3/_7\,{}^4/_7, {}^1/_7\,{}^4/_7, {}^4/_7\,{}^1/_7, {}^6/_7\,{}^3/_7$, and ${}^4/_7\,{}^3/_7$. These positions are indicated by solid circles in figure 4, which represents the proposed structure for the methyl ester polymer. One atom per residue has been placed in these positions; the calculated intensities correspond

to one oxygen atom and two carbon atoms for each asymmetric unit of three residues.

It is interesting to note that a reasonable explanation of the observed intensities of the equatorial reflections can be obtained with neglect of 40 per cent of the heavy atoms in the crystal. We believe that this phenomenon, which seems to be rather general for proteins and related substances, is due to the larger mutual interference of the rays scattered by the side-chain atoms than of those scattered by the main-chain atoms. The side-chain atoms, of which there are 90 in the unit, are located in 15 sets of 6-fold positions, presumably at 15 different values of the radius from the helical axis; whereas the 90 main-chain atoms (including the β carbon atoms) are more regularly arranged, corresponding to the pseudo 18-fold screw axis of the helix, and only five values of the radius are represented.

The anomalous reflections reported for the benzyl ester are described as lying on diffuse hyperbolas at 16 A and 8 A. We interpret these reflections as resulting from the presence of the 11-residue 3-turn helix, for which the value 16.8 A for c_0 would be predicted, on the assumption that the residue distance is the same as in the methyl ester, 1.53 A. This corresponds to 5.60 A per turn; Bamford, Hanby, and Happey suggested 5.76 A or a somewhat smaller value for the translation along the fiber axis, and we have found the data to correspond to $3 \times 5.76 = 17.3$ A, or 1.57 A per residue. The increase of 0.04 A over the methyl ester may well be due to van der Waals repulsion of the side chains.

The 11-residue 3-turn helix does not have a 6-fold axis or 3-fold axis, and accordingly it would not be expected to form hexagonal crystals with one helix per unit. The larger residue weight of the benzyl ester (219 in place of 143 for the methyl ester) would lead to a larger pseudo-hexagonal unit, with a_0 about 14.4 A, rather than 10.96 A (table 2). (The values given in table 2 are those obtained with the x-ray beam through the edge of an oriented film; closely similar values were also obtained with the x-ray beam perpendicular to the film, and with oriented fibers.)

Let us consider ways in which an asymmetric helix might be surrounded by others. We assume that the helixes are all equivalent, and that they are in contact with one another in a simple way. The six vectors from each asymmetric helix to its six neighbors are of six different kinds, representing six kinds of interaction. These might consist of three pairs, with the two of each pair differing only in polarity. It is found that there are only three arrangements of the three pairs in which the helixes are equivalent, and that none of them explains the occurrence of the 25-A reflection and the 14.4-A reflection for the benzyl polymer. It is possible, however, that two adjacent helixes interact in a way without polarity, to form a doublet. This would occur, for example, if one helix were related to the other by a 2-fold screw axis. There are many ways of arranging these doublets in

a crystal; one of the simplest, shown in figure 5, has monoclinic (pseudo-hexagonal) symmetry, and for the benzyl ester corresponds to axes a_0 = 25.0 A, b_0 = 17.3 A, c_0 = 14.42 A, and β = 90°. This unit accounts for all of the reflections observed for this polymer.

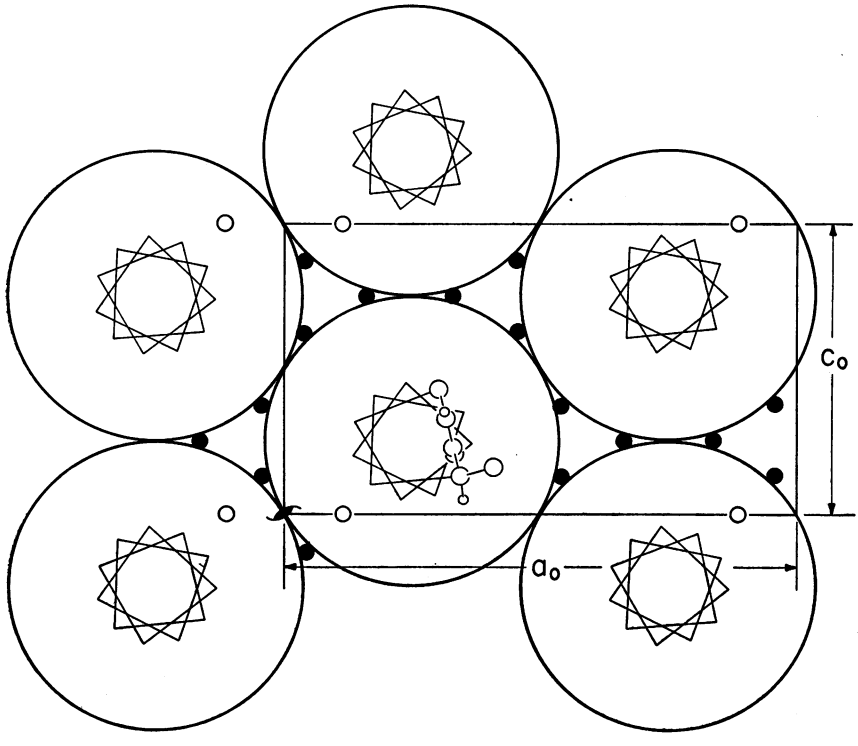

FIGURE 5

Proposed structure of poly-γ-benzyl-L-glutamate.

The intensities of all equatorial reflections can be roughly accounted for by consideration of only the five main-chain atoms, cylindrically distributed, as described above, about helical axes with coordinates x, z, and $\bar{x}$, $\bar{z}$, with $x = {}^1/_4$ and $z = {}^1/_4$, and one additional atom per residue, in the same positions as assumed for the methyl ester. The weak reflections $\{100\}$ and $\{001\}$ can be accounted for by any one of various slight distortions from the ideal hexagonal structure. The values of F_1 in table 2 are the calculated form factors for the five main-chain atoms per residue, and those of F_2 include the contributions of additional atoms in the positions assumed for the methyl ester (shown by full circles in figure 5). The scattering power in this position has been taken as 50 per cent greater than for

the methyl ester. An additional concentration of electrons equivalent to one oxygen atom per unit of eleven residues has also been placed at each of the positions x, $z = \pm 1/6$, 0 (shown by open circles in the figure). This small additional concentration of electrons, representing the deviation from the hexagonal symmetry of the methyl ester, accounts for the faint reflections $\{100\}$ and $\{001\}$. The close approximation of the structure to hexagonal packing of circular cylinders is made evident not only by the close approximation of the axial ratio a/c to $\sqrt{3}$, but also by the fact that the intensities of the two reflections requiring the larger unit can be explained by the scattering power of two heavy atoms, out of the 330 in the unit cell. Any one of many alternative small deviations of the distribution of scattering power from the ideal distribution with hexagonal symmetry would, of course, account for these observed weak reflections.

It seems likely that a poly-L-glutamate helix would be more stable with one screw sense (right-handed or left-handed) than with the other, and that helixes of only one kind are formed in significant number in the process of folding. A hexagonal array of helixes might then consist of equal numbers with positive and with negative orientation and essentially random distribution over the lattice points, equal numbers in an ordered array, only helixes with one orientation, or (much less likely) an ordered array in a ratio other than 1:1. For poly-γ-methyl-L-glutamate the detail of the photographs and the presence of only one helix per unit cell strongly suggest that the crystals contain helixes with only one orientation, and that segregation has occurred in crystallization. For the poly-benzyl ester the two-molecule unit also suggests that the helixes all have the same orientation; the strong reflections all have $h + k + l$ even, as required for an approximately body-centered unit. The possibility that segregation of positively and negatively oriented molecules occurs during crystallization suggests that annealing the specimens might improve the photographs.

Bamford, Hanby, and Happey also investigated several copolymers, and found from x-ray investigation that some tended to crystallize with the α configuration and some, especially those containing a large fraction of glycine residues, with the β configuration, involving sheets of extended polypeptide chains with lateral hydrogen bonds. They found that the α phase was oriented much more readily by stretching than the β phase, and attributed this difference to the presence of intramolecular hydrogen bonds in the α phase molecules. This feature remains unchanged by our attribution of our helical structure to the α phase. We explain the stability of the β phase for the glycine polymers by the stability of the lateral hydrogen bonds that can be formed; we have noted that with other residues there is serious steric interference between side chains, for the β configuration, leading to increase in the hydrogen-bond distance by 0.2 A, and a corresponding decrease in stability, relative to the α helix.

Bamford, Hanby, and Happey report interplanar distances leading to fiber-axis pseudo identity distances of 5.75 A for each of the three copolymers poly-DL-β-phenylalanine, poly-(DL-β-phenylalanine:γ-methyl-L-glutamate), and poly-(DL-β-phenylalanine:L-leucine). Larger steric repulsion of side chains would be expected for these copolymers than for poly-γ-methyl-L-glutamate.

The infrared dichroism observed by Ambrose and Elliott is accounted for very satisfactorily by our helical structure. In the 3.7-residue hydrogen-bonded helix the N—H and C=O bonds are oriented nearly parallel to the helical axis, the angle of deviation being only about 12°. This leads to a predicted dichroic ratio of about 44:1 for the N—H and C=O stretching vibrations. The largest observed dichroism, 14:1 for N—H stretching, is well within the predicted limit; it corresponds to an average angle of deviation of about 20° in the partially oriented specimen.

We conclude that there is strong evidence that crystals of poly-γ-methyl-L-glutamate and poly-γ-benzyl-L-glutamate, and also of the peptide copolymers, contain molecules with the third-amide hydrogen-bonded helical configuration, with about 3.7 residues per turn; and that the intermolecular forces operating in the crystals have caused the helixes in the first substance to assume the 18-residue 5-turn configuration, and those in the second to assume the 11-residue 3-turn configuration, these configurations having 3.60 and 3.67 residues per turn, respectively.

This investigation was aided by grants from The Rockefeller Foundation, The National Foundation for Infantile Paralysis, and The U. S. Public Health Service.

* Contribution No. 1551.

[1] Pauling, L., and Corey, R. B., *J. Am. Chem. Soc.*, **72**, 5349 (1950).

[2] Pauling, L., Corey, R. B., and Branson, H. R., these PROCEEDINGS, 205–211 (1951).

[3] Pauling, L., and Corey, R. B., *Ibid.*, to be published later.

[4] Bamford, C. H., Hanby, W. E., and Happey, F., *Proc. Roy. Soc.*, **A205**, 30 (1951).

[5] Ambrose, E. J., and Elliott, A., *Ibid.*, **A205**, 47 (1951).

[6] Huggins, M. L., *Chem. Rev.*, **32**, 195 (1943).

[7] Zahn, H., *Z. Naturforsch.*, **2B**, 104 (1947).

[8] Simanouti, T., and Mizushima, S., *Bull. Chem. Soc. Japan*, **21**, 1 (1948).

[9] Ambrose, E. J., and Hanby, W. E., *Nature*, **163**, 483 (1949).

[10] Astbury, W. T., *Ibid.*, **164**, 439 (1949).